

DOCUMENT RESUME

ED 052 245

TM 000 651

AUTHOR Bechtel, Gordon G.  
TITLE On Least Squares Fitting Nonlinear Submodels.  
INSTITUTION Educational Testing Service, Princeton, N.J.  
REPORT NO RB-71-11  
PUB DATE Mar 71  
NOTE 24p.  
  
EDRS PRICE MF-\$0.65 HC-\$3.29  
DESCRIPTORS Analysis of Variance, \*Mathematical Models,  
\*Mathematics, \*Multiple Regression Analysis,  
\*Statistics  
IDENTIFIERS \*Least Squares Method

ABSTRACT

Three simplifying conditions are given for obtaining least squares (LS) estimates for a nonlinear submodel of a linear model. If these are satisfied, and if the subset of nonlinear parameters may be LS fit to the corresponding LS estimates of the linear model, then one attains the desired LS estimates for the entire submodel. Two illustrative analyses employing this method are given, each involving an Eckart-Young (LS) decomposition of a matrix of linear LS estimates. In each case the factors provide an LS fit of the nonlinear submodel to the original data. The minimum error sum of squares for this fit is the error sum of squares for the corresponding linear model plus a function of the eigenvalues involved in the factorization. An Eckart-Young factorization, however, is only a special case of an LS decomposition of LS estimates. The present method is more generally applicable (under the three simplifying conditions) whenever any LS procedure may be found for fitting certain parameters of a nonlinear submodel to the corresponding LS estimates of a linear model. (Author)

RESEARCH

ED052245

ED052245

RB-71-11

U.S. DEPARTMENT OF HEALTH, EDUCATION  
& WELFARE  
OFFICE OF EDUCATION  
THIS DOCUMENT HAS BEEN REPRODUCED  
EXACTLY AS RECEIVED FROM THE PERSON OR  
ORGANIZATION ORIGINATING IT. POINTS OF  
VIEW OR OPINIONS STATED DO NOT NECES-  
SARILY REPRESENT OFFICIAL OFFICE OF EDU-  
CATION POSITION OR POLICY

# ON LEAST SQUARES FITTING NONLINEAR SUBMODELS

Gordon G. Bechtel  
Oregon Research Institute

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service  
Princeton, New Jersey  
March 1971

000 651

## ON LEAST SQUARES FITTING NONLINEAR SUBMODELS

Gordon G. Bechtel

Oregon Research Institute

### Abstract

Three simplifying conditions are given for obtaining least squares (LS) estimates for a nonlinear submodel of a linear model. If these are satisfied, and if the subset of nonlinear parameters may be LS fit to the corresponding LS estimates of the linear model, then one attains the desired LS estimates for the entire submodel. Two illustrative analyses employing this method are given, each involving an Eckart-Young (LS) decomposition of a matrix of linear LS estimates. In each case the factors provide an LS fit of the nonlinear submodel to the original data. The minimum error sum of squares for this fit is the error sum of squares for the corresponding linear model plus a function of the eigenvalues involved in the factorization. An Eckart-Young factorization, however, is only a special case of an LS decomposition of LS estimates. The present method is more generally applicable (under the three simplifying conditions) whenever any LS procedure may be found for fitting certain parameters of a nonlinear submodel to the corresponding LS estimates of a linear model.

# ON LEAST SQUARES FITTING NONLINEAR SUBMODELS<sup>1</sup>

Gordon G. Bechtel

Oregon Research Institute

Least squares estimation procedures are readily available for linear models which, depending upon their structure, generate "regression analyses" or "variance analyses" for particular data layouts. Since the parameters of nonlinear models are more difficult to estimate in an analytic way, their estimates are usually obtained by approximate iterative techniques. However, these techniques suffer from problems of local minima and may be rather unwieldy. Therefore, it is the purpose of this paper to present conditions which, if met by a particular nonlinear model, simplify the analytic problem of finding exact least squares estimates for its parameters.

Our approach is similar to that of the analysis of variance, where constraints (hypotheses) generate a linear submodel of a linear model. However, we shall use the general linear model as a device for embracing a nonlinear submodel of interest. This device makes it possible, under certain conditions, to obtain exact LS estimates for all of the submodel parameters by (1) LS fitting only the nonlinear parameters to the corresponding LS estimates of the linear model, and (2) preserving the remaining LS estimates of the linear model. In this way some difficult nonlinear least squares problems may be reduced to manageable form. Moreover, when the major interest is in the linear model itself, this technique constitutes an adjunct analysis, i.e., a further data breakdown, for the analysis of variance associated with that linear model.

The theoretical part of the paper presents the general linear model, and its nonlinear submodel, in partitioned form. This partitioning identifies

the subset of LS estimates to be submitted to further LS decomposition. It also permits the error sum of squares for the nonlinear submodel to be written as the (fixed and known) minimum error sum of squares for the linear model plus the sum of two other partitioned terms. These two terms represent additional error incurred under the nonlinear hypothesis, and this excess error sum of squares is analogous to the (linear) hypothesis sum of squares in the analysis of variance.

Subsequently, three conditions are invoked to reduce the sum of these two partitioned terms to a multiple of another error sum of squares, i.e., that for fitting the nonlinear parameters to the corresponding LS estimates of the linear model. The analytic minimization of this latter error sum of squares, if available, minimizes the entire error sum of squares for fitting the nonlinear submodel to the data. Two specific nonlinear problems, each satisfying these three conditions, and each amenable to an absolute minimization of the second error sum of squares, illustrate the usefulness of the method.

### Conceptual Approach

#### The Linear Model

Since the conceptual approach rests upon partitioned vectors and matrices, we write the general linear model as

$$(1) \quad y = (X_1'Z) \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + e,$$

where  $(X_1'Z)$  is the design matrix and  $\begin{pmatrix} \beta \\ \gamma \end{pmatrix}$  is the parametric vector. The model is fit to the vector  $y$  of observations with a resulting error vector  $e$ . A vector  $\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix}$  of least squares estimates is one which minimizes the error sum of squares  $e'e$  at  $\hat{e}'\hat{e}$ , where

$$\hat{e} = y - (X'Z) \begin{pmatrix} \hat{\beta} \\ -\hat{\gamma} \end{pmatrix}.$$

### The Nonlinear Submodel

The nonlinear model of interest may be generated by placing a hypothesis

$$(2) \quad H: \gamma = g(\alpha)$$

upon the linear model. The function  $g$  is a nonlinear constraint upon  $\gamma$  in the argument  $\alpha$ , which is some specified set of parameters. We then write the nonlinear submodel, i.e., the conjunction of (1) and (2), as

$$(3) \quad y = (X'Z) \begin{pmatrix} \beta \\ g(\alpha) \end{pmatrix} + f,$$

where  $f$  is a vector of errors. Our purpose in generating (3) as a submodel of (1) is not to test hypothesis (2), as is usually the case, but rather to use (2) as a device for LS estimation. This will be accomplished through the manipulation of the error sum of squares  $f'f$  for (3).

### The Error Sum of Squares for the Nonlinear Submodel

We may write the error vector for the submodel as

$$\begin{aligned} f &= y - (X'Z) \begin{pmatrix} \beta \\ g(\alpha) \end{pmatrix} \\ &= y - (X'Z) \begin{pmatrix} \hat{\beta} \\ -\hat{\gamma} \end{pmatrix} + (X'Z) \begin{pmatrix} \hat{\beta} \\ -\hat{\gamma} \end{pmatrix} - (X'Z) \begin{pmatrix} \beta \\ g(\alpha) \end{pmatrix} \\ &= \hat{e} + (X'Z) \begin{pmatrix} \hat{\beta} - \beta \\ -\hat{\gamma} - g(\alpha) \end{pmatrix}, \end{aligned}$$

where  $\epsilon = \hat{\gamma} - g(\alpha)$ . The least squares problem for (3) is that of choosing  $\beta, \alpha$  so as to minimize the error sum of squares

$$\begin{aligned} (4) \quad \mathcal{J}(\beta, \epsilon) &= f'f \\ &= \left( \hat{e} + (X'Z) \begin{pmatrix} \hat{\beta} - \beta \\ -\epsilon \end{pmatrix} \right)' \left( \hat{e} + (X'Z) \begin{pmatrix} \hat{\beta} - \beta \\ -\epsilon \end{pmatrix} \right) \\ &= \hat{e}'\hat{e} + 2(\hat{\beta}' - \beta' | \epsilon') \begin{pmatrix} -X'Z \\ \hat{\gamma} \end{pmatrix} \hat{e} + (\hat{\beta}' - \beta' | \epsilon') \begin{pmatrix} -X'Z \\ \hat{\gamma} \end{pmatrix} (X'Z) \begin{pmatrix} \hat{\beta} - \beta \\ -\epsilon \end{pmatrix}. \end{aligned}$$

Equation (4) shows that  $\mathcal{J}$  is the sum of a known, fixed value  $\hat{e}'\hat{e}$ , which is the error sum of squares for the corresponding linear model, and two other terms, which depend upon the submodel parameters  $\beta, \alpha$ .

#### LS Fitting $\alpha$ to $\hat{y}$

The following three paragraphs give conditions for obtaining an LS fit of  $\beta, \alpha$  to  $y$  by means of an LS fit of  $\alpha$  to  $\hat{y}$ :

Submodel invariance of  $\tilde{\beta}$ . The elements of  $\beta$  enter (3) linearly, while those of  $\alpha$  enter this submodel nonlinearly. In certain applications it may be possible to find the LS estimate  $\tilde{\beta}$  under the submodel, i.e., the value of  $\beta$  which minimizes  $\mathcal{J}$ , without simultaneously solving for the least squares estimate  $\tilde{\alpha}$ . Moreover, if

$$(5) \quad \tilde{\beta} = \hat{\beta},$$

which is the LS estimate under the corresponding linear model, then this value may be inserted into (4) reducing  $\mathcal{J}(\beta, \epsilon)$  to

$$(6) \quad \mathcal{J}'(\epsilon) = \hat{e}'\hat{e} + 2\hat{e}'Z\epsilon + \epsilon'Z'Z\epsilon.$$

$\mathcal{J}'$  is the sum of the error sum of squares for the corresponding linear model, a bilinear form in  $Z$ , and a quadratic form in  $Z'Z$ . Since these forms depend only upon  $\epsilon$ , and hence only upon  $\alpha$ , the LS solution for the nonlinear submodel is completed by choosing  $\alpha$  to minimize the sum of these bilinear and quadratic forms. Of course, this itself may present a formidable problem, and, therefore, we invoke two more conditions to reduce the latter sum to a multiple of  $\epsilon'\epsilon$ .

Vanishing bilinear form  $\hat{e}'Z\epsilon$ . If in a particular analysis the structure of  $\hat{e}$ ,  $Z$ , and  $\epsilon$  is such that

$$(7) \quad \hat{e}'Z\epsilon = 0,$$

then we have

$$(8) \quad \mathcal{J}'(\epsilon) = \hat{e}'\hat{e} + \epsilon'Z'Z\epsilon ,$$

which depends only upon  $\alpha$  through the quadratic form in  $Z'Z$ .

Reducible quadratic form  $\epsilon'Z'Z\epsilon$  . The structure of  $Z$  and  $\epsilon$  may also permit the reduction

$$(9) \quad \epsilon'Z'Z\epsilon = c\epsilon'\epsilon ,$$

of the quadratic form, where  $c$  is a known constant characteristic of the particular analysis. In this case, we can write

$$(10) \quad \mathcal{J}'(\epsilon) = \hat{e}'\hat{e} + c\epsilon'\epsilon ,$$

which depends only upon  $\alpha$  through the sum of squares  $\epsilon'\epsilon$ .

When conditions (5), (7), and (9) hold, the LS estimate  $\tilde{\alpha}$  may be found by minimizing  $\epsilon'\epsilon$ , which is the error sum of squares for the model

$$(11) \quad \hat{y} = g(\alpha) + \epsilon .$$

The LS problem for model (11) is, of course, simpler than that for model (3), which requires a direct minimization of  $\mathcal{J}$  in (4). The usefulness of (11) is illustrated in two nonlinear analyses in which (5), (7), and (9) are satisfied.

### Two Illustrative Analyses Involving Eckart-Young Decompositions

#### Gollob's Factor Analysis of Variance (FANOVA)

The linear model. The linear model for the standard two-way analysis of variance may be written in scalar form as

$$(12) \quad y_{ijk} = \mu + \theta_j + \eta_k + \rho_{jk} + e_{ijk} .$$

The  $\theta_j$  and  $\eta_k$  are main effects, the  $\rho_{jk}$  are interactions, and

$$(13a) \quad \sum_j \theta_j = \sum_k \eta_k = 0$$

$$(13b) \quad \sum_j \rho_{jk} = \sum_k \rho_{jk} = 0 \quad (j=1, \dots, J; k=1, \dots, K).$$

In the  $J \times K$  layout ( $J \geq K$ ) with  $m$  observations per cell ( $i=1, \dots, m$ ) the LS estimates are

$$(14) \quad \begin{aligned} \hat{\mu} &= y_{...} & \hat{\theta}_j &= y_{.j.} - y_{...} & \hat{\eta}_k &= y_{...k} - y_{...} \\ \hat{\rho}_{jk} &= y_{.jk} - y_{.j.} - y_{...k} + y_{...} \end{aligned}$$

where a dot indicates an average over the subscript it replaces. A typical element of the error vector  $\hat{e}$  is

$$\hat{e}_{ijk} = y_{ijk} - y_{.jk}$$

and the error sum of squares is

$$(15) \quad \hat{e}'\hat{e} = \sum_{ijk} (y_{ijk} - y_{.jk})^2.$$

The nonlinear submodel. Gollob's (1968) FANOVA model

$$(16) \quad y_{ijk} = \mu + \theta_j + \eta_k + \sum_{p=1}^r \lambda_{jp} \tau_{pk} + f_{ijk}$$

is a submodel of (12) which is formed by placing the interaction terms under the constraint

$$H: \rho_{jk} = \sum_{p=1}^r \lambda_{jp} \tau_{pk} \quad (j=1, \dots, J; k=1, \dots, K).$$

The hypothesis  $H$  sets each  $\rho_{jk}$  equal to an inner product over  $r (< K)$  dimensions. In (16) the  $\theta_j$  and  $\eta_k$  satisfy (13a), and the  $\lambda_{jp}$  and  $\tau_{pk}$  satisfy

$$(17a) \quad \sum_j \lambda_{jp} = \sum_k \tau_{pk} = 0 \quad (p=1, \dots, r),$$

$$(17b) \quad \sum_k \tau_{pk}^2 = 1 \quad (p=1, \dots, r),$$

$$(17c) \quad \sum_j \lambda_{jp} \lambda_{jq} = \sum_k \tau_{pk} \tau_{qk} = 0 \quad (p \neq q) .$$

Equation (17a) replaces (13b), while (17b) and the orthogonality conditions (17c) are required for the uniqueness of the  $\lambda_{jp}$  and  $\tau_{pk}$  under the factorization H.

Identifications. In examining the FANOVA model we make the following identifications with the preceding notation:

$$\beta \equiv \{\mu, \theta_j, \eta_k\} ,$$

$$\gamma \equiv \{\rho_{jk}\} ,$$

$$\epsilon \equiv \{\epsilon_{jk}\} ,$$

where

$$\epsilon_{jk} \equiv \hat{\rho}_{jk} - \sum_p \lambda_{jp} \tau_{pk} .$$

In the two-way layout with  $m$  observations per cell  $Z$  may be partitioned vertically, and the submatrix  $Z_i$  for the  $i^{\text{th}}$  replicate is given in Table 1. The rows and columns are labeled by the observations  $y_{ijk}$  and the interactions  $\rho_{jk}$ . Since this submatrix, which is the  $JK \times JK$  identity matrix, is repeated for each replicate  $i=1, \dots, m$ , the matrix  $Z$  may be written as

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_i \\ \vdots \\ Z_m \end{pmatrix} = \begin{pmatrix} I_{JK} \\ \vdots \\ I_{JK} \\ \vdots \\ I_{JK} \end{pmatrix} .$$

-----  
Insert Table 1 about here  
-----

Submodel invariance of  $\{\hat{\mu}, \hat{\theta}_j, \hat{\eta}_k\}$ . The error sum of squares for (16) is

$$\mathcal{S} = \sum_{ijk} f_{ijk}^2,$$

and, introducing Lagrange multipliers  $\phi$  and  $\psi$  for (13a), we generate

$$\mathcal{M} = \mathcal{S} + \phi \sum_j \theta_j + \psi \sum_k \eta_k.$$

Setting to zero the derivatives of  $\mathcal{M}$  with respect to  $\mu$ ,  $\theta_j$ , and  $\eta_k$ , using (13a) and (17a), and eliminating  $\phi$  and  $\psi$ , we find that

$$\begin{aligned} \tilde{\mu} &= \hat{\mu}, \\ (18) \quad \tilde{\theta}_j &= \hat{\theta}_j, \\ \tilde{\eta}_k &= \hat{\eta}_k, \end{aligned}$$

which are given by (14).

The vanishing bilinear form. The bilinear form in this analysis is

$$\hat{\epsilon}' Z \epsilon = \sum_{i,j,k} \sum_{l,m} \hat{e}_{ijk} z_{ijk,lm} \epsilon_{lm},$$

where the element of  $Z$  is

$$z_{ijk,lm} = \begin{cases} 1 & \text{when } l,m = j,k \\ 0 & \text{otherwise} \end{cases}.$$

Thus we have

$$\begin{aligned} (19) \quad \hat{\epsilon}' Z \epsilon &= \sum_{ijk} \hat{e}_{ijk} \epsilon_{jk} \\ &= \sum_{jk} \epsilon_{jk} \sum_i \hat{e}_{ijk} \\ &= 0, \end{aligned}$$

since

$$\sum_i \hat{e}_{ijk} = \sum_i (y_{ijk} - y_{.jk}) = 0 \quad .$$

The reduction of the quadratic form. The matrix  $Z'Z$  is

$$\begin{aligned} Z'Z &= \sum_{i=1}^m Z_i'Z_i \\ &= mI_{JK} \quad . \end{aligned}$$

Therefore, the quadratic form may be written as

$$\begin{aligned} (20) \quad \epsilon'Z'Z\epsilon &= \epsilon'mI_{JK}\epsilon \\ &= m\epsilon'\epsilon \\ &= m \sum_{jk} \epsilon_{jk}^2 \quad , \end{aligned}$$

and the error sum of squares for (16) is

$$(21) \quad \mathcal{S}' = \sum_i \sum_j \sum_k (y_{ijk} - y_{.jk})^2 + m \sum_{jk} \epsilon_{jk}^2 \quad .$$

The Eckart-Young decomposition of the  $\hat{\rho}_{jk}$ . Due to (21) we may obtain LS estimates of the  $\lambda_{jp}$  and  $\tau_{pk}$  by LS fitting these parameters to the  $\hat{\rho}_{jk}$  through the model

$$\hat{\rho}_{jk} = \sum_p \lambda_{jp} \tau_{pk} + \epsilon_{jk} \quad ,$$

which, in matrix notation, is

$$(22) \quad \hat{P} = \Lambda T + E \quad ,$$

where

$\hat{P} = (\hat{\rho}_{jk})$  is the  $J \times K$  matrix of estimated interactions,  
 $\Lambda = (\lambda_{jp})$  is a  $J \times r$  matrix ,

$T = (\tau_{pk})$  is an  $r \times K$  matrix,

$E = (\epsilon_{jk})$  is a  $J \times K$  matrix of residual errors.

The LS subproblem for (22) is easily solved by an Eckart-Young decomposition of  $\hat{P}$ . This provides estimates  $\tilde{\Lambda}$  and  $\tilde{T}$  which minimize

$$\epsilon' \epsilon = \sum_{j,k} \epsilon_{jk}^2,$$

and hence the second term of  $\mathcal{S}$  in (21). The LS estimate of  $T$  in (22) is given by

$$(23) \quad \tilde{T} = (\tilde{\tau}_{pk}) = \begin{pmatrix} \tilde{\tau}_1 \\ \vdots \\ \tilde{\tau}_r \end{pmatrix},$$

where the rows  $\tilde{\tau}_1, \dots, \tilde{\tau}_r$  are  $r$  orthonormal (unit length) eigenvectors corresponding to the  $r (\leq K - 1)$  largest eigenvalues of the  $K \times K$ , rank  $K - 1$  matrix  $\hat{P}'\hat{P}$  (e.g., see Eckart & Young, 1936; Householder & Young, 1938; Keller, 1962; Whittle, 1952). With  $\tilde{T}$  thus constructed, the LS estimate of  $\Lambda$  is obtained as

$$(24) \quad \tilde{\Lambda} = (\tilde{\lambda}_{jp}) = \hat{P}'\tilde{T}.$$

In this particular least squares solution  $\tilde{T}$  is row-wise orthonormal, and  $\tilde{\Lambda}$  is column-wise orthogonal, i.e.,

$$(25a) \quad \tilde{T}\tilde{T}' = I,$$

$$(25b) \quad \tilde{\Lambda}'\tilde{\Lambda} = L,$$

where the  $r$  diagonal elements

$$\ell_1 \geq \dots \geq \ell_r > 0$$

of the diagonal matrix  $L$  are the  $r$  largest eigenvalues of  $\hat{P}'\hat{P}$ . Equations (25a) and (25b) restate (17b) and (17c) in matrix notation.

The Eckart-Young decomposition for (22) completes the solution since it provides the  $\tilde{\lambda}_{jp}$  and  $\tilde{\tau}_{pk}$  which, along with  $\tilde{\mu}_i$ ,  $\tilde{\theta}_j$ , and  $\tilde{\eta}_k$ , are exact LS estimates for fitting (16) to the original data  $y_{ijk}$ . The associated minimum error sum of squares for (16) may be found by noting that

$$\begin{aligned}
 (26) \quad \min \sum_{jk} \sum \epsilon_{jk}^2 &= \sum_{jk} \tilde{\epsilon}_{jk}^2 \\
 &= m \sum_{p=r+1}^{K-1} \ell_p,
 \end{aligned}$$

where the latter summation is taken over the  $(K-1)-r$  positive eigenvalues not used in the construction of  $\tilde{T}$  (Keller, 1962). Since we have chosen the rows of  $\tilde{T}$  as  $r$  orthonormal eigenvectors corresponding to the  $r$  largest eigenvalues of  $\hat{P}\hat{P}$ , this summation is over the  $(K-1)-r$  smallest positive eigenvalues. Due to (26) the minimum of (21) is

$$\begin{aligned}
 (27) \quad \min \mathcal{S} &= \hat{e}'\hat{e} + c \min \epsilon'\epsilon \\
 &= \sum_{ijk} (y_{ijk} - y_{.jk})^2 + m \sum_{p=r+1}^{K-1} \ell_p,
 \end{aligned}$$

which is the error sum of squares in LS fitting the FANOVA submodel to the replicated two-way layout. Expression (27) consists of the usual error sum of squares for this layout incremented by a second term representing the  $(K-1)-r$  deleted factors in the Eckart-Young decomposition of the rank  $(K-1)$  matrix of estimated interactions. Of course, if all  $K-1$  factors are extracted, then the second term vanishes, and the error sum of squares takes a minimum value equal to that for the two-way analysis of variance.

#### The Multidimensional Choice Scaling of Bechtel, Tucker, & Chang

The linear model. In a three-way layout of graded paired comparisons a typical observation  $y_{ijk}$  indicates the strength of choice for object  $j$  over object  $k$  in replicate  $i$ , e.g., individual  $i$  (Bechtel, Tucker & Chang,

in press). When  $n$  objects are being studied, each replicate contains  $\binom{n}{2}$  comparisons, and hence there are  $m\binom{n}{2}$  observations in a design involving  $m$  ( $\geq n$ ) replicates. The linear model for this pairwise choice layout is

$$(28) \quad y_{ijk} = \theta_{ij} - \theta_{ik} + \rho_{jk} + e_{ijk},$$

where the  $\theta_{ij}$  and  $\theta_{ik}$  are intrareplicate scale values, the  $\rho_{jk}$  are inter-replicate interactions ( $\rho_{jk} = -\rho_{kj}$ ), and

$$(29a) \quad \sum_j \theta_{ij} = 0 \quad (i=1, \dots, m)$$

$$(29b) \quad \sum_{k \neq j} \rho_{jk} = 0 \quad (j=1, \dots, n).$$

The LS estimates of the parameters of (28) are

$$(30) \quad \begin{aligned} \hat{\theta}_{ij} &= \frac{1}{n} \sum_{k \neq j} y_{ijk}, \\ \hat{\rho}_{jk} &= y_{.jk} - \hat{\theta}_{.j} + \hat{\theta}_{.k}, \end{aligned}$$

a typical element of the error vector  $\hat{e}$  is

$$\hat{e}_{ijk} = y_{ijk} - \hat{\theta}_{ij} + \hat{\theta}_{ik} - \hat{\rho}_{jk},$$

and the error sum of squares for fitting the model is

$$(31) \quad \hat{e}'\hat{e} = \sum_{i,j} \sum_{k \neq j} (y_{ijk} - \hat{\theta}_{ij} + \hat{\theta}_{ik} - \hat{\rho}_{jk})^2.$$

The nonlinear submodel. The observational equation

$$(32) \quad y_{ijk} = \sum_{p=1}^r \lambda_{ip} \tau_{pj} - \sum_{p=1}^r \lambda_{ip} \tau_{pk} + \rho_{jk} + f_{ijk}.$$

given by Bechtel, Tucker, and Chang (in press), is generated by placing the constraint

$$H: \theta_{ij} = \sum_{p=1}^r \lambda_{ip} \tau_{pj} \quad (i=1, \dots, m; \quad j=1, \dots, n)$$

upon the scale values in (28). In (32) the  $\rho_{jk}$  satisfy (29b), and the  $\lambda_{ip}$  and  $\tau_{pj}$  satisfy

$$(33a) \quad \sum_j \tau_{pj} = 0 \quad (p=1, \dots, r) ,$$

$$(33b) \quad \sum_j \tau_{pj}^2 = 1 \quad (p=1, \dots, r) ,$$

$$(33c) \quad \sum_j \tau_{pj} \tau_{qj} = \sum_i \lambda_{ip} \lambda_{iq} = 0 \quad (p \neq q) .$$

Equation (33a) replaces (29a), while (33b) and the orthogonality conditions (33c) are required for the uniqueness of the  $\lambda_{ip}$  and  $\tau_{pj}$  under the factorization H.

Identifications. The following definitions link the components of the choice model to the preceding general notation:

$$\beta \equiv \{\rho_{jk}\} ,$$

$$\gamma \equiv \{\theta_{ij}\} ,$$

$$\epsilon \equiv \{\epsilon_{ij}\} ,$$

where

$$\epsilon_{ij} \equiv \hat{\theta}_{ij} - \sum_p \lambda_{ip} \tau_{pj} .$$

In this replicated paired comparisons design the matrix Z, which is  $m \binom{n}{2} \times mn$ , may be partitioned into  $m^2$  submatrices, each being  $\binom{n}{2} \times n$ . The submatrix  $Z_{ii'}$ , for replicates  $i$  and  $i'$ , is the zero matrix for  $i \neq i'$ . For  $i = i'$  it is the pairs  $\times$  singles scale matrix in Table 2, where the rows and columns are labeled by the observations  $y_{ijk}$  and the scale values  $\theta_{ij}$ . Since this submatrix is invariant over replicates  $i = 1, \dots, m$ , the matrix Z may be written as

$$Z = \begin{pmatrix} Z_{11} & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & Z_{ii} & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & Z_{nn} \end{pmatrix} = \begin{pmatrix} S & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & S & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & S \end{pmatrix}.$$

-----  
 Insert Table 2 about here  
 -----

Submodel invariance of  $\{\tilde{\rho}_{jk}\}$ . The error sum of squares for (32) is

$$\mathcal{S} = \sum_{i,j} \sum_{k} f_{ijk}^2,$$

and, introducing Lagrange multipliers  $\phi_j$  for (29b), we have

$$\mathcal{M} = \mathcal{S} + \sum_j \phi_j \sum_{k \neq j} \rho_{jk}.$$

Setting to zero the derivative of  $\mathcal{M}$  with respect to  $\rho_{jk}$ , using (29b) and (33a), and eliminating the  $\phi_j$ , we find that

$$(34) \quad \tilde{\rho}_{jk} = \hat{\rho}_{jk},$$

which is given by (30).

The vanishing bilinear form. The bilinear form in the error sum of squares for this analysis is

$$\hat{e}' Z e = \sum_{i,j,k} \hat{e}_{ijk} \sum_{l,m} \hat{e}_{ijk} z_{ijk,lm} e_{lm},$$

where the element of  $Z$  is

$$z_{ijk,lm} = \begin{cases} 1 & \text{when } l,m = i,j \\ -1 & \text{when } l,m = i,k \\ 0 & \text{otherwise} \end{cases}.$$

Thus we have

$$\begin{aligned}\hat{e}'Z\epsilon &= \sum_{ij} \sum_k \hat{e}_{ijk} \epsilon_{ij} - \sum_{ij} \sum_k \hat{e}_{ijk} \epsilon_{ik} \\ &= \sum_{ij} \sum_k \hat{e}_{ijk} (\epsilon_{ij} - \epsilon_{ik}) ,\end{aligned}$$

and, since  $\hat{e}_{ijk} = -\hat{e}_{ikj}$ , we may write

$$\begin{aligned}(35) \quad \hat{e}'Z\epsilon &= \frac{1}{2} \sum_{ij} \sum_{k \neq j} \hat{e}_{ijk} (\epsilon_{ij} - \epsilon_{ik}) \\ &= \frac{1}{2} \sum_{ij} \epsilon_{ij} \sum_{k \neq j} \hat{e}_{ijk} - \frac{1}{2} \sum_{ik} \epsilon_{ik} \sum_{j \neq k} \hat{e}_{ijk} \\ &= 0 .\end{aligned}$$

Equation (35) follows from the fact that

$$\sum_{j \neq k} \hat{e}_{ijk} = \sum_{k \neq j} \hat{e}_{ijk} = \sum_{k \neq j} (y_{ijk} - \hat{\theta}_{ij} + \hat{\theta}_{ik} - \hat{\rho}_{jk}) = 0$$

due to (29a), (29b), (30), and the skew-symmetry of  $\{\hat{e}_{ijk}\}$ .

The reduction of the quadratic form. The matrix  $Z'Z$  for (28) is

$$Z'Z = \begin{pmatrix} S'S & . & . & . & 0 & . & . & . & 0 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ 0 & . & . & S'S & . & . & . & . & 0 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ 0 & . & . & . & 0 & . & . & S'S \end{pmatrix}$$

where

$$S'S = \begin{pmatrix} n-1 & \dots & -1 & \dots & -1 \\ \vdots & & \vdots & & \vdots \\ -1 & \dots & n-1 & \dots & -1 \\ \vdots & & \vdots & & \vdots \\ -1 & \dots & -1 & \dots & n-1 \end{pmatrix}.$$

Therefore, the quadratic form in the error sum of squares for (32) is

$$\begin{aligned} \epsilon'Z'Z\epsilon &= \begin{pmatrix} \epsilon'_1 & \dots & \epsilon'_i & \dots & \epsilon'_m \end{pmatrix} \begin{pmatrix} S'S & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & S'S & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & S'S \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_m \end{pmatrix} \\ &= \sum_{i=1}^m \epsilon'_i S'S \epsilon_i, \end{aligned}$$

where  $\epsilon'_i$  is the  $i^{\text{th}}$  row of the matrix  $(\epsilon_{ij})$ . Letting  $(1)$  denote an  $n \times n$  matrix with each entry equal 1, we may reduce the  $i^{\text{th}}$  quadratic form under the latter summation to

$$\begin{aligned} \epsilon'_i S'S \epsilon_i &= \epsilon'_i [nI - (1)] \epsilon_i \\ &= \epsilon'_i nI \epsilon_i - \epsilon'_i (1) \epsilon_i \\ &= n \epsilon'_i \epsilon_i, \end{aligned}$$

which follows from the fact that  $\epsilon'_i(1)$  is the zero vector, i.e.,

$$\sum_j \epsilon_{ij} = \sum_j (\hat{\theta}_{ij} - \sum_p \lambda_{ip} \tau_{pj}) = 0 \quad (i=1, \dots, m)$$

due to (29a) and (33a). Using this reduction we then write the entire quadratic form as

$$(36) \quad \begin{aligned} \epsilon'Z'Z\epsilon &= n \sum_{i=1}^m \epsilon_i' \epsilon_i \\ &= n \sum_{ij} \epsilon_{ij}^2, \end{aligned}$$

and hence the error sum of squares for (32) as

$$(37) \quad \mathcal{S}' = \sum_{ij} \sum_{k} (y_{ijk} - \hat{\theta}_{ij} + \hat{\theta}_{ik} - \hat{\rho}_{jk})^2 + n \sum_{ij} \epsilon_{ij}^2.$$

The Eckart-Young decomposition of the  $\hat{\theta}_{ij}$ . Equation (37) enables us to obtain LS estimates of the  $\lambda_{ip}$  and  $\tau_{pj}$  by LS fitting these parameters to the  $\hat{\theta}_{ij}$  through the model

$$\hat{\theta}_{ij} = \sum_p \lambda_{ip} \tau_{pj} + \epsilon_{ij},$$

which, in matrix notation, is

$$(38) \quad \hat{\Theta} = \Lambda T + E,$$

where

$$\begin{aligned} \hat{\Theta} &= (\theta_{ij}) \text{ is the } m \times n \text{ matrix of estimated scale values,} \\ \Lambda &= (\lambda_{ip}) \text{ is an } m \times r \text{ matrix,} \\ T &= (\tau_{pj}) \text{ is an } r \times n \text{ matrix,} \\ E &= (\epsilon_{ij}) \text{ is an } m \times n \text{ matrix of residual errors.} \end{aligned}$$

An Eckart-Young decomposition of  $\hat{\Theta}$  in (38) provides estimates  $\tilde{\Lambda}$  and  $\tilde{T}$  which minimize

$$\epsilon' \epsilon = \sum_{ij} \epsilon_{ij}^2,$$

and hence the second term of  $\mathcal{S}'$  in (37). The construction of  $\tilde{\Lambda}$  and  $\tilde{T}$ ,

which has been described in detail by Bechtel, Tucker and Chang (in press), is similar to the construction of the estimated factors in the FANOVA model.

The Eckart-Young decomposition for (38) completes the analysis by providing the  $\tilde{\lambda}_{ip}$  and  $\tilde{\tau}_{pj}$  which, along with the  $\tilde{\rho}_{jk}$ , are LS estimates for fitting (32) to the original paired comparisons  $y_{ijk}$ . The error sum of squares for this fit is the minimum of  $\mathcal{S}'$  in (37) and is attained at the minimum

$$(39) \quad n \sum_{ij} \tilde{\epsilon}_{ij}^2 = n \sum_{p=r+1}^{n-1} \ell_p,$$

of the second term of  $\mathcal{S}'$ . Since  $\text{rank } \hat{\theta} = \text{rank } \hat{\theta}'\hat{\theta} = n-1$  due to (29a), the summation on the right is over the  $(n-1)-r$  smallest positive eigenvalues of  $\hat{\theta}'\hat{\theta}$ . Substituting (39) into (37) gives

$$(40) \quad \min \mathcal{S}' = \hat{e}'\hat{e} + c \min \epsilon'\epsilon$$

$$= \sum_{ij} \sum_{k} (y_{ijk} - \hat{\theta}_{ij} + \hat{\theta}_{ik} - \hat{\rho}_{jk})^2 + n \sum_{p=r+1}^{n-1} \ell_p,$$

which is the error sum of squares for LS fitting the choice submodel to the three-way layout of paired comparisons. In (40) the error sum of squares for the corresponding linear model is incremented by a second term whenever the number  $r$  of factors in the Eckart-Young decomposition is less than the rank  $(n-1)$  of the matrix of estimated scale values. When  $r = n-1$  the second term vanishes, and the error sum of squares is minimized at the first term.

### References

- Bechtel, G. G., Tucker, L. R., & Chang, W. An inner product model for the multidimensional scaling of choice. Psychometrika, in press.
- Eckart, C., & Young, G. The approximation of one matrix by another of lower rank. Psychometrika, 1936, 1, 211-218.
- Gollob, H. F. A statistical model which combines features of factor analytic and analysis-of-variance techniques. Psychometrika, 1968, 33, 73-115.
- Householder, A. S., & Young, G. Matrix approximation and latent roots. American Mathematical Monthly, 1938, 45, 165-171.
- Keller, J. B. Factorization of matrices by least-squares. Biometrika, 1962, 49, 239-242.
- Whittle, P. On principal components and least square methods of factor analysis. Skandinavisk Aktuarietidskrift, 1952, 35, 223-239.

Footnote

The developments in this paper were initiated and carried out in part while the author was a Visiting Research Fellow at Educational Testing Service. They were extended and concluded at the Oregon Research Institute under Grant Nos. MH 12972 and MH 15506 from the National Institute of Mental Health, U. S. Public Health Service.

TABLE 1

The  $i$ th submatrix  $Z_i = I_{JK}$

	$\rho_{11}$	$\dots$	$\rho_{jk}$	$\dots$	$\rho_{JK}$
$y_{i11}$	1	$\dots$	0	$\dots$	0
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$y_{ijk}$	0	$\dots$	1	$\dots$	0
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$y_{iJK}$	0	$\dots$	0	$\dots$	1

TABLE 2  
The Pairs  $\times$  Singles Scale Matrix  $Z_{ji} = S$

	$\theta_{i1}$	$\theta_{i2}$	...	$\theta_{ij}$	...	$\theta_{ik}$	...	$\theta_{i,n-1}$	$\theta_{in}$
$y_{i12}$	1	-1	...	0	...	0	...	0	0
.	.	.		.		.		.	.
.	.	.		.		.		.	.
.	.	.		.		.		.	.
$y_{ijk}$	0	0	...	1	...	-1	...	0	0
.	.	.		.		.		.	.
.	.	.		.		.		.	.
.	.	.		.		.		.	.
$y_{i,n-1,n}$	0	0	...	0	...	0	...	1	-1